Janus Vision and Roadmap

Prepared By:

Houtan Aghili, PhD
Avijit Chatterjee, PhD
IBM Software Group

Elise Blaese, MS, MBA
Kathleen Martin, MS
IBM Global Business Services

Kamal Narang, MBA
CTIS

November 3, 2006

TABLE OF CONTENTS

1. Executive Summary

Janus is a standards-based clinical data repository that utilizes an open source data model. The data model for Janus was designed jointly in 2004 through a Cooperative Research and Development Agreement (CRADA) between the Federal Drug Administration (FDA) and IBM. Janus provides a data collection and analysis repository for clinical and non-clinical (animal toxicology) trial data submitted for protocols as well as clinical outcomes data. This data is submitted to the FDA following the Study Data Tabulation Model (SDTM) and Standard for Exchange of Nonclinical Data (SEND) developed by the Clinical Data Interchange Standards Consortium (CDISC).

The use of CDISC standards as the foundation of Janus is important for its ability to be used as a platform for collaboration in the clinical development ecosystem. The CDISC standards are developed with representation from the international regulatory community as well as biotechnology and pharmaceutical (Biopharmaceutical) sponsors. This has enabled the vision of an effective data interchange environment for submission and review of drug and biologics applications. The vision is further augmented by the Biomedical Research Integrated Domain Group (BRIDG) semantic model, whose development is supported both by CDISC and by the cancer Biomedical Informatics Grid (caBIG™) initiative sponsored by the National Cancer Institute (NCI).

This document was developed to provide a roadmap for the progression of Janus over the next 24 months at the FDA and NCI, and to outline Janus' impact to the overall clinical development ecosystem over the next five years. As part of this effort, we have interviewed over 30 key individuals in the clinical development ecosystem (FDA, NCI, academia, and industry) to learn how Janus can evolve to meet their needs and improve the clinical development process.

Janus facilitates cross-trial analysis for a broad range of clinical and research users. With a syntactically and semantically robust content, Janus enables users to perform data mining tasks across completed trials for the development of future clinical studies, and to compare findings and adverse events against certain criteria. Janus also enables the FDA's review of Investigational New Drugs (INDs), New Drug Applications (NDAs), and Biologic License Applications (BLAs).

Improving both accuracy and timeliness of drug safety and efficacy decisions is currently a high-priority issue for consumers, regulatory reviewers, pharmaceutical and biotechnology industries. With the realization of the Janus vision, a comprehensive repository of clinical trial data will be available to regulatory reviewers, drug developers, and researchers, enabling them to conduct complex standards-based cross-trial analyses. As a result, they will be able to identify anomalies and outliers as part of their normal workflow much more effectively than at present. This will result in the faster approval of safer, more efficacious therapies.

Janus interfaces with commercial analytical tools that ultimately will integrate using an "open toolbox" (proposed by FDA), which would allow interoperability and reusability of analysis modules between the tools. With the help of the "open toolbox" supported on top of a common data repository Janus would enhance sharing of analyses across the FDA, NCI, and their biopharmaceutical and healthcare partners. The FDA and NCI have already made their commitments to foster more coordinated and efficient clinical research clear in their respective vision documents [R1, R2].

There is significant pressure on the key players in drug development and clinical trial communities to adopt the emerging data interchange standards, common vocabulary approaches, and the required semantic technology for linking the disparate sources where today's knowledge resides. This document outlines this vision, focusing on the short-term and long-term rollout within specific FDA and NCI divisions, departments and centers. It also addresses how to encourage the adoption of SDTM and Janus within the biopharmaceutical industry in the absence of a FDA regulation for adoption of CDISC standards.

2. Positioning

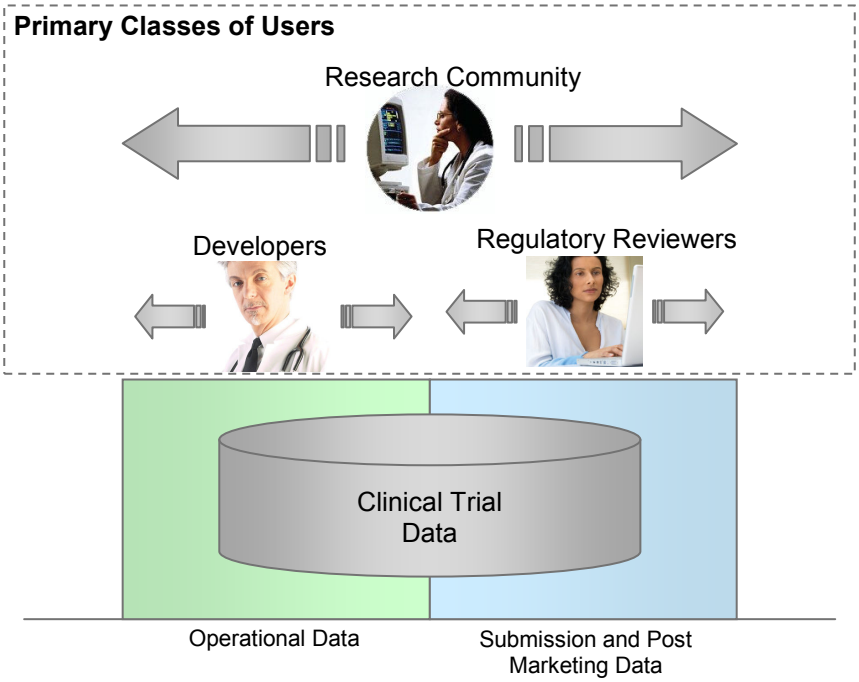
2.1. Vision statement

The overarching vision of the Janus project is

1. To develop a knowledge base capturing the entire life cycle of clinical research data and to make this knowledge base available to a diverse community of users with interest in regulatory review, research and analysis.
2. To develop a collaboration framework that fosters communications amongst this diverse set of users, expediting the approval process and enabling effective collaborative research and development.
3. To develop a solution to leverage open standards that would facilitate interoperability within an enterprise and amongst distinct institutions built based on modern software engineering paradigms, and in particular Service Oriented Architecture (SOA) [R3].

We define the life cycle of clinical research data in a broad term covering its operational phases as well as after its submission to the FDA and post-approval marketing. This broad scope implies that a unified solution with the above vision in mind should accommodate a wide spectrum of data models, usage scenarios, and access patterns. As such we can classify the user constituency into the following three general classes as depicted by Figure 1:

1. Reviewers at the FDA with interest in regulatory review aspects of the data, who are primarily interested in clinical trial data once it is submitted to the FDA.
2. Researchers at the NCI, FDA, academic medical centers and biopharmaceutical companies, who have interest in clinical research and would like to see both the operational data and the summarized data once it is submitted to the FDA.
3. The development community whose interest is primarily the operational data for clinical development purposes.



Clinical Trial Data Life Cycle

Figure 1. Solution Scope

All of the above constituents should be able to explore cross clinical trial data analysis and not limit their scope to a single study at a time, which is the common silo model at the moment.

Once a common platform with syntactically and semantically robust content is established for data analysis and review, users will be in a position to collaborate much more effectively than they can with the current tedious and awkward process. Currently, those engaged in analysis in the research community and during regulatory reviews capture results in data files. They then augment this data with text documents describing how the results were obtained and share this information with another party via regular mail and e-mail, hoping that the other party can reproduce the results.

With the common analysis platform in place, as shown in Figure 2, the description of a study will be captured electronically, as part of research or review setup, and expressed in a standard formal language. This will include the specification of the assumptions and preconditions, the input data sets, the review and analysis steps, and the analysis and review results. The end work product, an "analysis capsule", can then be shared electronically with other interested parties as part of an inter-enterprise collaboration workflow. The recipients will use the "analysis capsule" to reproduce the original analysis.

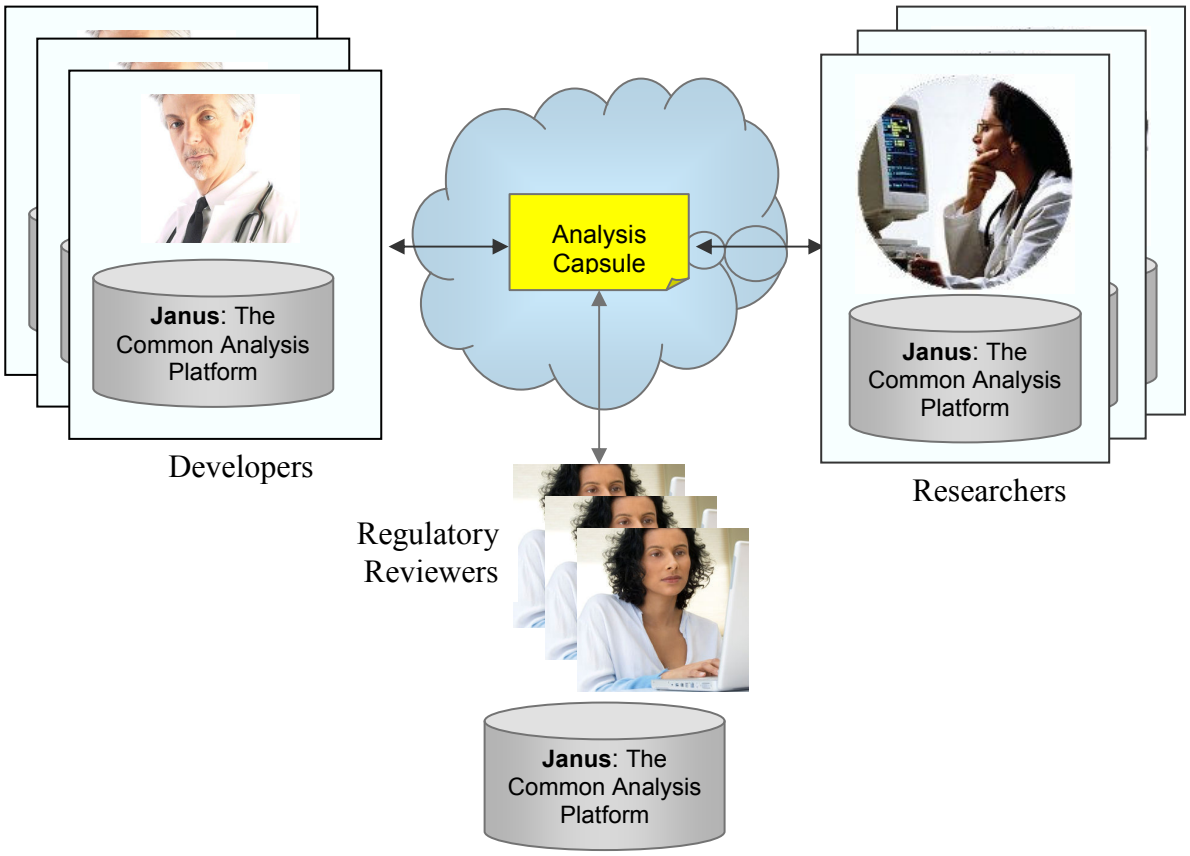


Figure 2. Inter-Enterprise Collaboration Framework

Although the vision of Janus is to address the comprehensive life cycle of clinical trial data, we will restrict our attention to regulatory review and analysis. There are other current initiatives aimed at the operational data, which will complement the portion of the vision that is described here.

2.2. Problem & solution statement

Any robust solution must be anchored in a set of current relevant activities that have already paved the way for the development of such solutions. Moreover, the interoperability requirements necessitate leveraging the applicable open standards. Figure 3 is a high-level depiction of how we envision applying some of the existing efforts and open standards.

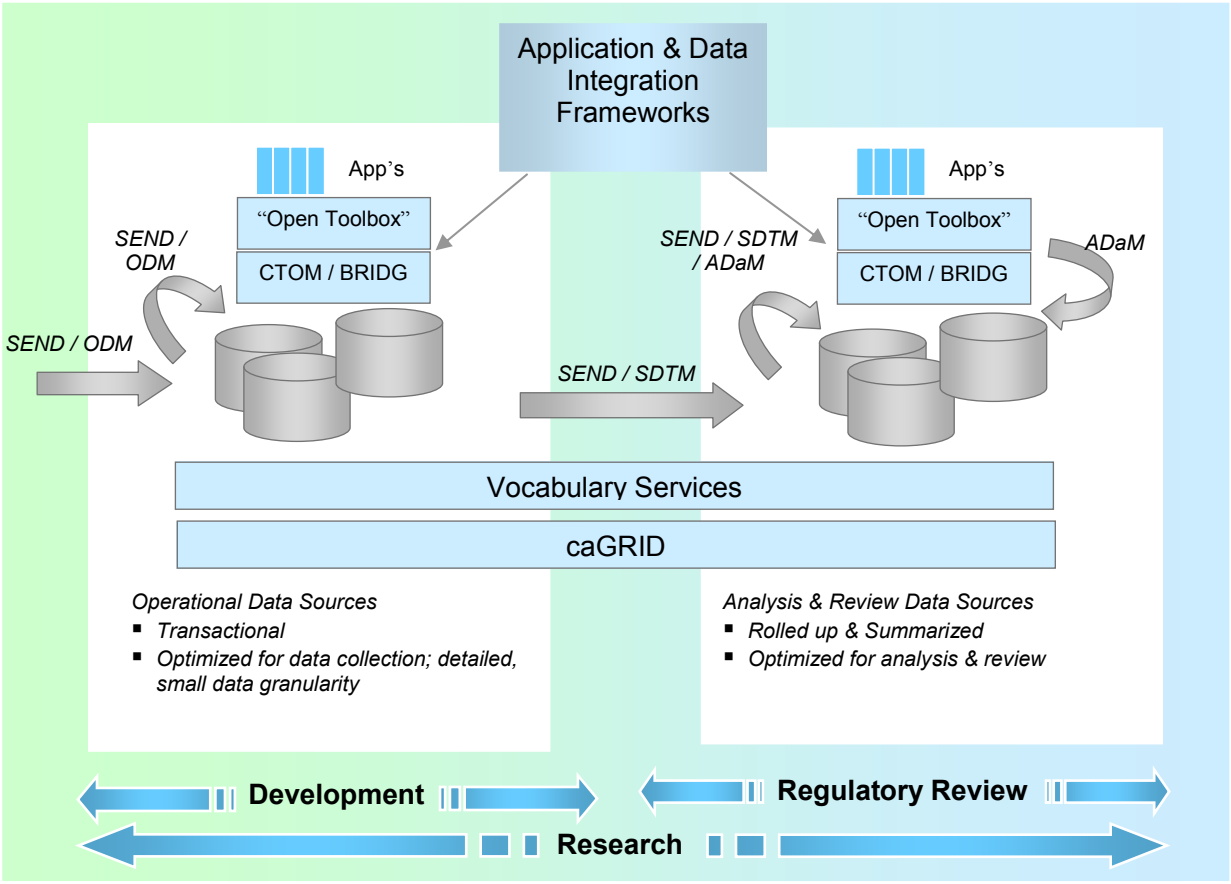


Figure 3. Janus Solution in Relation to Existing Initiatives

As indicated, the vision of Janus leverages the following:

1. The standards being developed by the CDISC Standards body:
    a. Operational Data Model (ODM) for the exchange of operational data.
    b. Study Data Tabulation Model (SDTM) for the exchange of clinical data.
    c. Standard for Exchange of Nonclinical Data (SEND) for the exchange of non-clinical data.
    d. Analysis Data Model (ADaM) for capturing analysis models, steps, and results, enabling effective collaboration amongst users.
2. The standards being developed as part of NCI's caBIG™ initiative, and in particular:
    a. The Clinical Trial Object Model (CTOM; see Section 4.2) and BRIDG for standard clinical trial data syntax and capturing its semantics (a common controlled Lexicon) [R4, R5]. We consider CTOM/BRIDG as the means for providing a data integration layer that address both syntactic and semantic aspects of the data. Providing semantically harmonized data is a *prerequisite* for enabling the objective comparison of observations and thus robust cross-trial analysis.
    b. caGRID standard for enabling connectivity amongst multiple repositories across multiple enterprises [R6].

c. NCI's Enterprise Vocabulary Services (EVS) will provide a comprehensive library of controlled vocabulary relevant for coding clinical trial and non-clinical data [R7]. CDISC is building a controlled vocabulary for SDTM terms which will be maintained in EVS. Janus has an entity called mapping table that allows mapping between different controlled vocabularies being used for coding the same term.

3. The Open Toolbox initiative being launched by the FDA provides a common framework for introducing analytical and review tools so that they can be shared and utilized easily amongst the user community. Here analysis is performed as a choreographed flow of granule, reusable, (globally) shared analysis steps. An analyses scenario is then captured in scripts that produced the analysis datasets, with well-defined input and output sources, rather than storing the analysis datasets themselves. This will significantly reduce the barriers for sharing the analysis results amongst users.

Again as stated earlier, we will restrict our attention to regulatory review and analysis, i.e., the right-hand side of Figure 3.

2.3. Opportunity and Benefit

Improving Safety and Efficacy: With the implementation of the Janus vision, a standardized and comprehensive repository of clinical trial data will be available to regulatory reviewers, drug developers, and researchers, enabling them to conduct complex cross-trial analyses. As a result, they will be able to discover anomalies and outliers as part of their normal workflow much more effectively than currently, where in many instances they are limited to single-study analysis. This will result in the faster approval of safer, more efficacious therapies.

Accelerating time to market: Another significant concern of consumers and drug developers alike is the time to market for new medications. The current laborious approval process requires significant communication between regulatory reviewers and sponsors across very inefficient channels, and is typically plagued with considerable time-consuming paper-work. Realizing the Janus vision, establishing a high-bandwidth and effective collaboration platform between stakeholders (*i.e.*, regulatory reviewers, sponsors, and researchers) will enable the development and approval processes to be conducted significantly faster than the current practice.

Reducing Development Risk: Under the current practice, there have been several instances of late stage failures in new drug candidates, resulting in significant unnecessary development expenditure. Janus offers an analysis platform which could be used to avoid such occurrences by identifying ineffective agents earlier in the development life cycle, and stop or modify the trial design.

3. Stakeholder Community

The key stakeholders who will implement and interact with the Janus clinical data repository are:

1. **Food & Drug Administration (FDA)**
   - Reviews and approves IND / BLA and NDA submissions
2. **Sponsors**
   a. **National Cancer Institute (NCI)**
      - Performs, facilitates, and coordinates clinical trials
      - Conducts clinical / basic research
      - Submits INDs to, and collaborate with FDA for review and approval
      - Submits periodic reports (safety, efficacy, etc.) to government agencies (Congress, etc.)

   b. **Biopharmaceuticals Companies and Academic Medical Centers**
      - Conduct, facilitate, and coordinate clinical trials
      - Submit INDs / BLAs / NDAs to, and collaborate with FDA for review and approval

3. **Clinical Research Information Exchange (CRIX)**
   - Host Federal Investigator Registry for Biomedical Informatics Research Data (FIREBIRD) service automating the FDA Form 1572 registration process
   - Provide SDTM and Janus data validation service to the industry to help them prepare for the electronic submission of an IND, NDA or BLA application.
   - Offer an open and SOA based analysis toolbox framework that would support interoperability of analysis procedures across a variety of review and analytical products running on top of Janus

Figure 4 illustrates the envisioned data flow into and out of Janus for FDA, NCI and Industry. Each stakeholder could have its own instance of Janus. They would communicate between each other using Study Data Tabulation Model (SDTM), Standard for Exchange of Nonclinical Data (SEND) and Analysis Dataset Model (ADaM) data interchange standards. The respective Janus instances at NCI and Biopharmaceutical sites will also receive derived data from analysis datasets in ADaM format.[1]

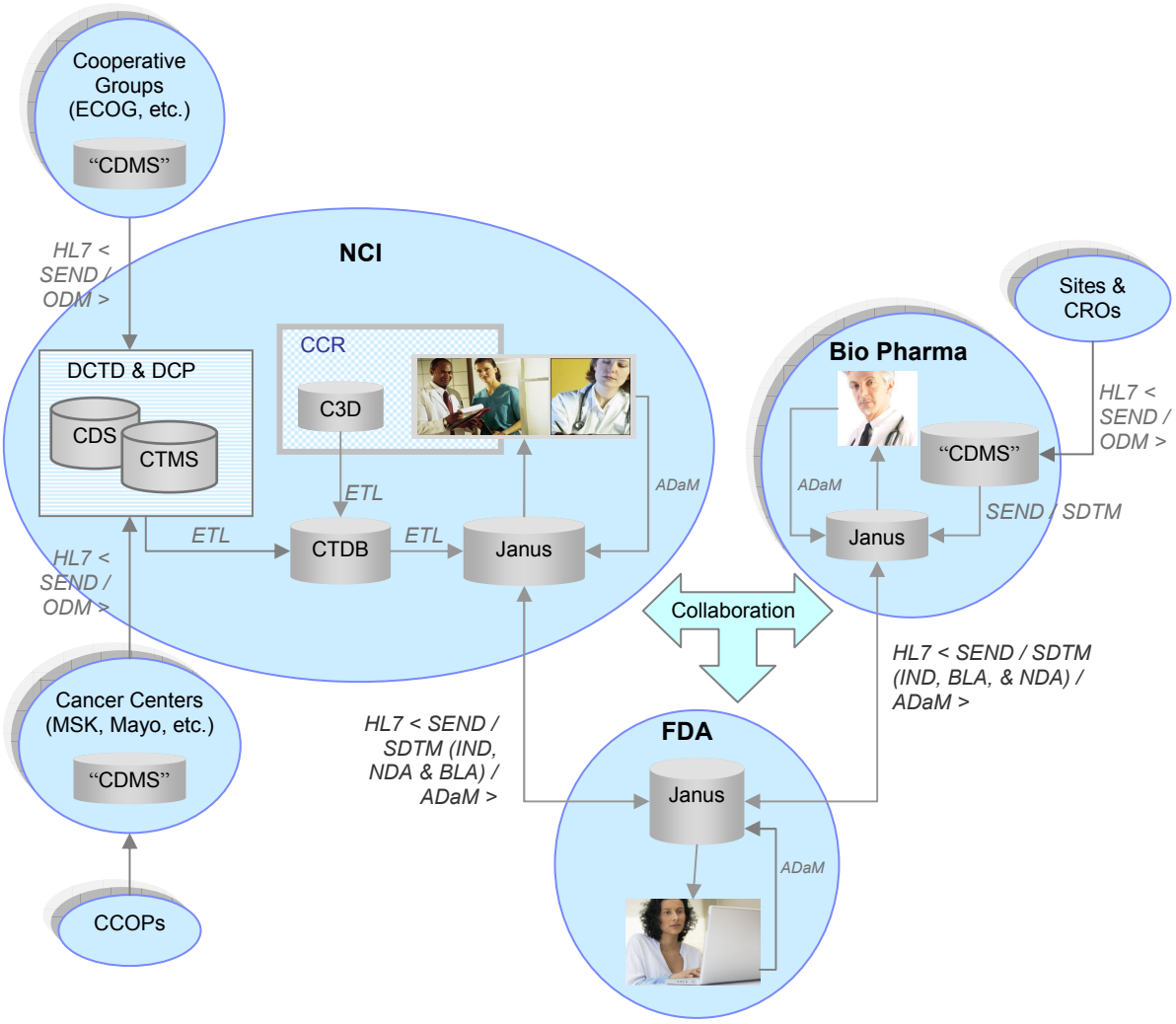


Figure 4. Envisioned data flow between Janus clinical data repository stakeholders

[1] As indicated earlier, Janus currently does not support the storage of analysis data sets but will do so as the CDISC finalizes the specification of the ADaM data standard.

The NCI has recently undergone a comprehensive evaluation of its clinical trials process under the auspices of the Clinical Trial Working Group (CTWG). The CTWG's proposed vision is the storage of all NCI operational trial data in a centralized repository called the Clinical Trial Data Base (CTDB). The data feeding CTDB will come from trials sponsored or funded by Cancer Therapy Evaluation Program (CTEP) and Division of Cancer Prevention (DCP) via Cancer Centers, Cooperative Groups and Community Clinical Oncology Programs (CCOPs). Besides CTDB, operational clinical trial data from the NCI's Center for Cancer Research (CCR) stored in their Cancer Central Clinical Database (C3D) will also be converted into SDTM and loaded into NCI's instance of Janus.

For biopharmaceutical companies and academic medical centers that make IND, NDA or BLA submissions to FDA, operational data are obtained from the sites and Contract Research Organizations (CROs) that execute their clinical trials. The operational clinical trial data (including Case Report Form and Laboratory data) and non-clinical animal toxicology data will be stored in the respective Clinical Data Management System (CDMS) repositories within the Biopharmaceutical enterprises from which SDTM and SEND data will be extracted and loaded into their local Janus instances.

The profiles of the key member groups for each of the four stakeholder constituents FDA, NCI, Biopharmaceutical and CRIX are summarized in the following Sections.

3.1. FDA

3.1.1. Clinical Reviewers

Representative	Center for Drug Evaluation and Research (CDER)
Description	Clinical Medical Officers from the from Office of Biostatistics, Office of Translational Science
Type	MDs specialized clinical background and strong understanding of disease mechanism and drug action
Responsibilities	Review of IND/NDA clinical data. Conduct both pre- and post-marketing evaluation of new drugs for safety and efficacy, and provide guidance to IND/NDA sponsors.
Success Criteria	Capability to provide access to SDTM data stored in Janus via review applications such as I-Review, SAS, WebSDM, S-Plus and PPV.
Involvement	Key decision makers in the review of NDA and BLA applications
Deliverables	Prepare decision letters and various interim communication memos with sponsors during the review of an application
Comments/Issues	Currently lack the facility for easy querying of submission datasets per application to prepare for analysis

3.1.2. Statisticians

Representative	CDER
Description	Statisticians from the Division of Biometrics
Type	Strong statistical background
Responsibilities	Evaluate the adequacy of study designs and statistical analysis plans for proposed clinical studies during the IND process. Evaluate study conclusions drawn from the study drug efficacy and safety data during the NDA review. Statistical reviews also may be requested for evaluation of drug carcinogenicity and stability data.
Success Criteria	Capability to provide access to SDTM and SEND data stored in Janus via statistical applications such as SAS, S-Plus and store derived data and analysis results back into Janus
Involvement	Assist in the review of IND, NDA and BLA applications
Deliverables	Create analysis datasets for reviewers and do detailed statistical analysis
Comments/Issues	Cannot do sophisticated meta-analysis on cross-trial datasets very readily

3.1.3. Pharmacology Reviewers

Representative	CDER
Description	Pharmacologist/toxicologists are assigned to review divisions within the Office of New Drugs, CDER
Type	Mostly with Ph.D. and Post-graduate degrees in Pharmacology
Responsibilities	Evaluate laboratory animal and *in vitro* (preclinical) studies that have been conducted with a drug — this includes data from *in vitro* and *in vivo* studies that determine the drug's pharmacological effects and toxicological profile. Review the effects on fertility, pregnancy and fetal development, carcinogenicity, safety pharmacology, genotoxicity, and those related to the absorption, distribution, metabolism and excretion data.
Success Criteria	Capability to provide access to SEND data stored in Janus via review applications such as ToxVision
Involvement	Key decision makers in the review of IND applications
Deliverables	Prepare decision letter and various interim communication memos with sponsors during the review of an application
Comments/Issues	Mostly review spreadsheet datasets, but would like to review a richer submission in SEND format assuming data is easily accessible from Janus via tools such as ToxVision

3.1.4. Biopharmacist / Clinical Pharmacologist

Representative	CDER
Description	From the Office of Clinical Pharmacology, Office of Translational Science, CDER
Type	Mostly with Ph.D. and Post-graduate degrees in Pharmacology
Responsibilities	Evaluate and interpret biopharmaceutic and clinical pharmacology information in INDs/NDAs that have an impact on efficacy and safety of medical treatments

3.2. *NCI*

3.2.1. Cancer Therapy Evaluation Program (CTEP)

Representative	Cancer Centers, Cooperative Groups and Community Clinical Oncology Program
Description	Sponsor and/or fund cancer trials where NCI may or may not own the investigational drug candidate
Type	Large number of sites nationwide that are funded under this program
Responsibilities	Improve lives of cancer patients by finding better ways to treat, control and cure cancer
Success Criteria	Provide cross-trial clinical research capability for cancer researchers via Janus
Involvement	May submit the IND applications to FDA directly or assist sponsors on their submission
Deliverables	Collect operational data in CTDB that will feed into Janus
Comments/Issues	Currently Cooperative groups only submit partial data (demographic) into CDUS and Phase 1 complete data in CTMS and thus CTDB will need to be comprehensive for complete operational data for all trial phases so that richer data can be stored in Janus

3.2.2. Division of Cancer Prevention (DCP)

Representative	Research groups divided into Foundations of Prevention (Basic Prevention Science, Biometry, Cancer Biomarkers, Chemopreventive Agent Development, Community Oncology and Prevention Trials, Early Detection and Nutritional Science) and Organ Systems (Breast and Gynecologic Cancer, Gastrointestinal and Other Cancer, Lung and Upper Aerodigestive Cancer, Prostate and Urologic Cancer)
Description	Funds trials for prevention and detection of specific cancers and study lifestyle issues diet, food and nutrition and substance use (alcohol, tobacco and drugs)
Type	The respective research groups consist of basic and clinical scientists who support a variety of cancer research grants, contracts and program-initiated research
Responsibilities	Promote cancer cure by doing basic and clinical research on novel cancer prevention and detection methods
Success Criteria	Provide cross-trial research capability for the research groups
Involvement	May not submit applications to FDA but provide basic research reports on cancer prevention and detection to the cancer research community
Deliverables	Store all operational data in CTDB which will then feed into Janus
Comments/Issues	Can benefit from correlating Demographic and Subject Characteristic domain data with those in the findings and events domain which will all be stored in Janus

3.2.3. Center for Cancer Research (CCR)

Representative	Conduct basic research on new investigational drugs for cancer at CCR
Description	Conduct clinical trials in-house at CCR to study safety issues on new investigational drugs
Type	World renowned clinical researchers in the field of cancer with MD and Ph.D degrees
Responsibilities	Study toxicity of new investigational drugs in cancer
Success Criteria	Provide cross-trial clinical research capability and assist in designing new protocols and patient recruitment
Involvement	Submit IND applications to FDA
Deliverables	Store operational data in C3D that will then be loaded into Janus after de-identification
Comments/Issues	Can significantly benefit from cross-trial analysis of de-identified C3D and CTDB data all loaded into Janus

3.2.4. Division of Extramural Activities

Representative	Fund R03, R21 grants for NCI and Trans-NIH Initiatives to external academic cancer research centers
Description	Promotes innovative research of cutting edge initiatives such as Genes and the Environment, Molecular Markers, Cancer Imaging, Defining Signatures and others.
Type	Clinical and Basic Researchers in prestigious academic research institutions
Responsibilities	Conduct innovative and cutting-edge research on cancer
Success Criteria	Provide cross-trial clinical research capability to design new experiments for innovative research
Involvement	Does not interact with FDA but submits research reports to NCI for review
Deliverables	Generate annual reports and status updates on the studies funded by the Grants
Comments/Issues	Eventually some of this data can be loaded into CTDB and then perhaps to Janus if relevant for cross-trial analysis

3.2.5. Cancer Trials Support Unit (CTSU)

Representative	NCI Pilot Project supporting a national network of oncologists
Description	Assists oncologists across the nation to enroll their patients in Phase-III cancer treatment trials sponsored by the NCI Cooperative Groups
Type	Physicians across the nation sign up their patients for these trials
Responsibilities	Provide latest cancer treatment options to a wider patient community
Success Criteria	Simple portals such as PDQ need to be provided for detailed trial information to the oncologists to assist them in picking the right trial for their patients
Involvement	The NDAs are usually submitted by the Cooperative Groups or by the Sponsors manufacturing the cancer drug involved in the trial.
Deliverables	The Cooperative Groups currently provide summary data to NCI and store operational data in their own internal CDMS repositories.
Comments/Issues	Would be nice to standardize ODM data export from the cooperative group CDMS repositories to load into CTDB which would then feed into Janus

3.3. Biopharmaceutical & Academic Medical Centers

3.3.1. Biotechnology and Pharmaceutical Companies

Representative	Pharmaceutical and Biotechnology companies
Description	Discover, develop and manufacture new drug candidates (small molecules or biologics) and submit for approval to FDA
Type	For profit institutions that are either public or private
Responsibilities	Execute animal studies for IND and human clinical trials (Phase 1, 2 and 3) for NDA and BLA applications
Success Criteria	Collaborate with FDA for the review of their IND, NDA and BLA applications via common data platform Janus
Involvement	Submit IND, NDA and BLA applications to FDA and obtain approval before drug can be marketed for a particular disease indication
Deliverables	Submit SDTM, SEND data to FDA as part of eCTD electronic submissions
Comments/Issues	Current process of collaboration for a drug application review is very tedious and time consuming. Janus with a common tool interface will significantly speed up the review process and allow efficient sharing of analysis results and observations between biopharmaceuticals and FDA

3.3.2. Academic Medical Centers

Representative	Academic Medical Centers such as Mayo, MSK that submits its own investigational drug applications
Description	Discover and develop new drug candidates (small molecules or biologics) and submit for approval to FDA and then license the approved drug to biopharmaceutical industry for manufacturing
Type	Prestigious research institutions providing cutting research medical care to patients using new and existing drugs
Responsibilities	Execute animal studies for IND and human clinical trials (Phase 1, 2 and 3) for NDA and BLA applications
Success Criteria	Collaborate with FDA for the review of their IND, NDA and BLA applications via common data platform Janus
Involvement	Usually submit IND applications and then co-submits NDA and BLA applications with a partner in the biopharmaceutical industry that manufactures the drug
Deliverables	Submit SDTM, SEND data to FDA as part of eCTD electronic submissions
Comments/Issues	Current process of collaboration for a drug application review is very tedious and time consuming. Janus with a common tool interface will significantly speed up the review process and allow efficient sharing of analysis results and observations.

3.4. *CRIX*

Representative	Members from biopharmaceutical Industry, Academic Research Institutions, FDA and NCI, and CDISC
Description	Provide services to stakeholders to enable new data sharing standards
Type	Independent organization setup with funding from biopharmaceutical and additional governing members from NCI, FDA and Academic sites.
Responsibilities	Provide services for Investigator Registration and Credentialing system called Firebird and support SDTM/Janus data validation services and promote a common tools interface for enhanced collaboration between FDA, NCI and Industry.
Success Criteria	Assist stakeholders to adopt new data sharing standards such SDTM, SEND and submitting data into Firebird
Involvement	Will act as a services organization for FDA and NCI to help implement business policies and validations on submission data and manage investigator data registry system (Firebird)
Deliverables	Provide the stakeholders access to the Firebird investigator registration database, support common tools infrastructure for analyzing data loaded in Janus and deliver validation services for data interchanges such as SDTM and SEND between FDA, NCI and Industry
Comments/Issues	This is a trusted third-party to be setup with industry funding to enable enhanced collaboration between FDA, NCI and Biopharmaceutical

4. Project Scope Overview

4.1. Description

FDA has been very proactive in promoting electronic regulatory submissions from sponsors. FDA has also worked with the industry through the CDISC standards body to develop the SDTM standard for the tabulation datasets and ADaM standard for the analysis datasets. The goal of the standards is to improve the productivity of the review process by speeding up the data preparation step and promote stronger collaboration among reviewers at FDA and between FDA and sponsors. This would be accomplished by using standardized review tools to access, view, manipulate and analyze the tabulation datasets. These standards are also expected to benefit the industry in streamlining the flow of clinical data from collection through submission and facilitating data interchange between providers and partners.

The Janus relational data repository was designed to store tabulation datasets such as SDTM for human trials and SEND for animal toxicology data. Tabulation datasets, which comprise individual observations for a subject (collected in a clinical trial) or animals (collected in toxicology experiments), are one of the four domains of data submitted by a sponsor to FDA as part of an IND, NDA or a BLA application. The other domains are patient profiles, listings and analysis files. SDTM 3.1.1, the latest version of the standard, however, does not fully support analysis data files, and therefore, the sponsors still need to submit these along with the tabulation files. CDISC is developing the ADaM standard for analysis data files.

The data model for Janus was designed using the artifacts of the SDTM design paradigm, primarily intended for efficient storage, query, review, and reporting of SDTM and SEND tabulation datasets. Janus uses a normalized design, reducing the domains into the classes which are then mapped into respective entities. Furthermore, Janus supports an extensible schema through the Entity-Attribute-Value (EAV) design of the Comments and Qualifiers tables that allow mapping of new attributes added into the SDTM and SEND domains.

For complex and study-specific analysis and data mining, “analysis-ready” data marts may be created from a Janus database. For example, materialized SDTM and SEND views may be created on top of Janus, enabling efficient access using review and analysis tools that can consume SDTM and SEND data. Observe, however, that Janus currently does not support the storage of analysis data sets, but

will do so as the CDISC finalizes the specification of the ADaM data standard. See Appendix 7.1 for more details on the Janus model.

Figure 5 shows the data flow from the operational repository through Janus into the analytical data marts. As indicated, data is exported in SDTM and SEND format from the operational data stores first into Janus (for efficient storage, query, review and reporting) and then again into analytical data marts for complex study-specific data safety and efficacy analysis scenarios; materialized SDTM / SEND views may be regarded as the simplest examples of such "analysis ready" repositories.

The operational store could be any Clinical Data Management System (CDMS), for instance:

- For FDA: the Electronic Document Room (EDR) that stores all the incoming eCTD submissions containing the tabulation datasets in Modules 4 and 5;
- For NCI: the Clinical Trial Database (CTDB) proposed by CTWG that would capture data on all cancer trials pulled from systems such as CDS, CTMS, C3D and others at the cancer centers, cooperative groups and extramural programs; and
- For Biopharmaceuticals: Oracle Clinical, Phase Forward's ClinTrial or any other custom warehouse that stores CRF and Lab data.

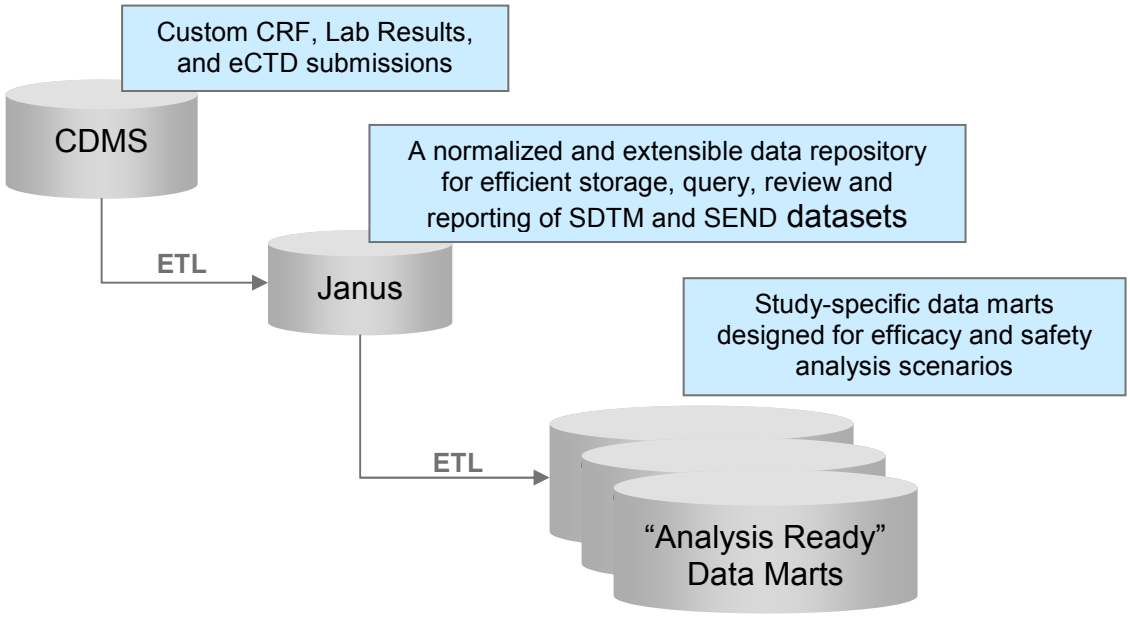


Figure 5. Data flow from operational to submission to analytical data repositories

4.2. Reference Architecture

Figure 6 depicts the reference architecture for the Janus solution. As indicated, we envision clinical trial content being extracted from operational sources, validated and transformed into the proper format, and then loaded into a Janus data repository. The operation of these steps will be driven by the currently accepted specifications from standards bodies (e.g., CDISC SDTM/SEND), and any relevant business rules (e.g., FDA's rules for accepting submission data).

The Janus data repository will then serve as the base for the generation of analysis-specific data marts, e.g., specific analysis-ready data marts intended for specific safety and efficacy scenarios (which will be incrementally updated every time a new study is added into Janus). The collection of Janus and data marts, together with other internal data sources (image repositories, genomics data, etc.) and external data sets (GenBank, PubMed, etc.) provide the analytic data platform for regulatory review, development, research and analysis. The data federation layer provides an optimized means for accessing the data stored in the various data sets that are available for review and analysis. It will create a dynamic federated warehouse over a heterogeneous set of data sources, e.g., data marts

and external reference data sets like GenBank and PubMed. Without a data federation layer, the burden of integrating data in various sources will be placed on the application logic, replicated in each and every application. This would potentially add significantly to the complexity of the overall solution.

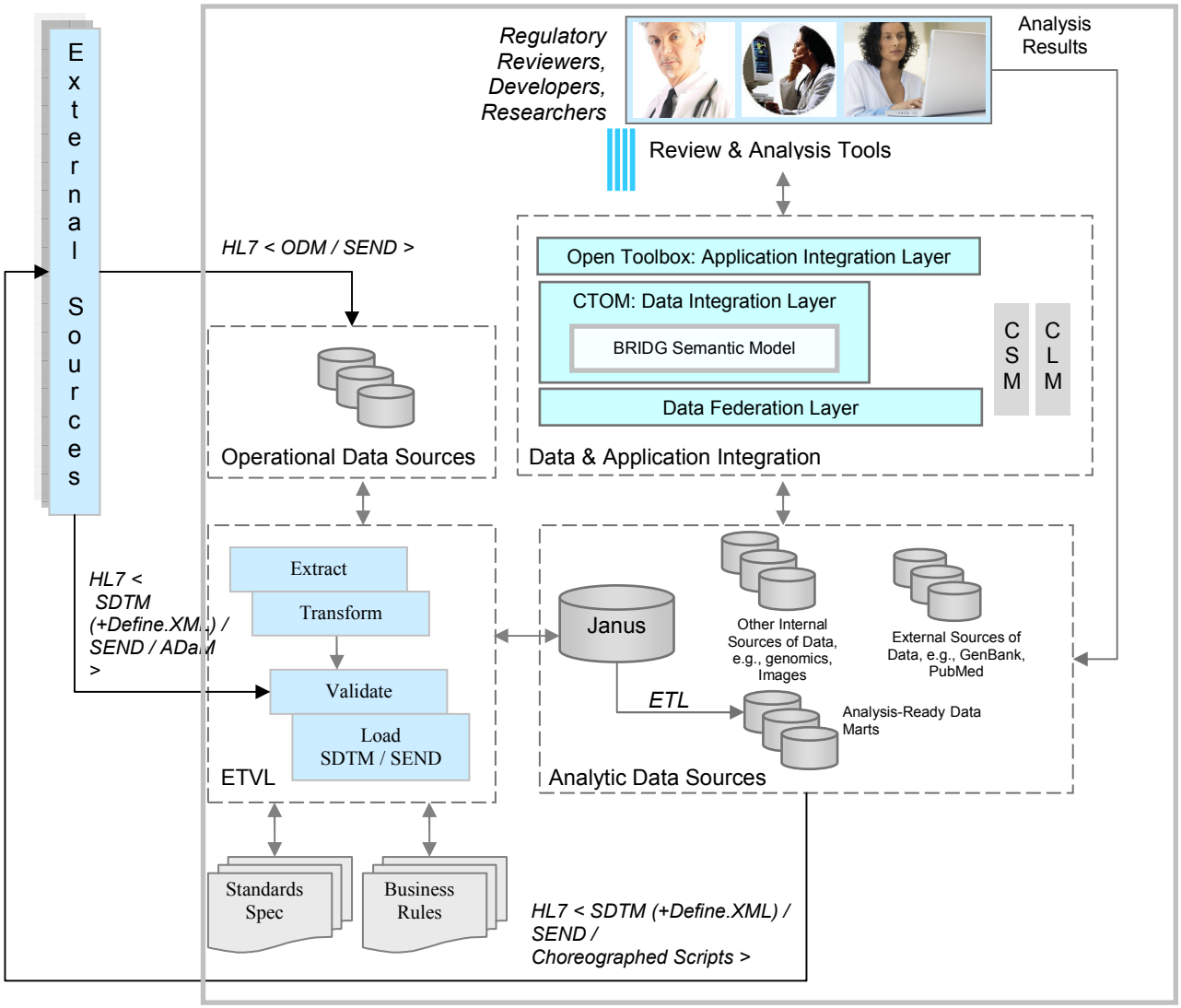


Figure 6. Janus Solution Reference Architecture

At the next level of abstraction, we envision leveraging NCI's caBIG™ vision to provide a storage independent object model, the Clinical Trial Object Model (CTOM) based on the BRIDG semantic model. NCI's caBIG initiative has proposed CTOM as the standard access layer to relational analytical data marts. We envision CTOM being the layer providing the semantic data integration platform. Observe, however, that some data sources may be accessed via other access layers through the federation layer – a data mart may be accessed via an SQL interface, for example.

NCI's caBIG™ initiative has also developed the Common Security Module (CSM) for enforcing user authentication and access control authorization in a uniform and consistent fashion (in conjunction with other enterprise directory services like LDAP, etc.). In addition, NCI has developed the Common Logging Module (CLM) for enforcing required accountability and audit trial in a regulated environment [R7]. These aspects are particularly important, because any deployment of a Janus solution will

include a suite of review and analysis tools developed by independent vendors, making the uniform enforcement data and applications security and auditability a challenge.

Finally, we envision leveraging FDA's "Open Toolbox" proposal as the layer that provides application integration. This will support reusability and interoperability of analysis modules amongst analytical applications.

As indicated, an enterprise may exchange data with external entities:

- Operational clinical trial data is exchanged using ODM and SEND via HL7 transport;
- Summarized data will be exchanged using SDTM and SEND via HL7 transport; and
- Analysis data will be exchanged using ADaM via HL7 transport.

4.3. Services Oriented Architecture

A Service Oriented Architecture (SOA) decomposes a solution into a set of services. A service is "an encapsulation of a software function" which supports the execution of a repeatable business task. Within the SOA programming model, developers build services and use services to compose solutions. Services are built into components. Components present the service interfaces that are used by other solution components to exchange information and request work.

A services orientation views a business as a set of linked services and the outcomes that these services produce. In this regard, SOA represents the Component Business Modeling approach, which deconstructs a business into business components, which in turn provide services to the rest of the enterprise. An SOA services oriented architecture is therefore an IT architectural style that supports service orientation.

By designing a solution according to the concepts and principles of a SOA, architects can more readily identify services and components that must be specialized to selective usage, as opposed to services and components that are more common. Common services can be intentionally designed for reuse. An SOA helps designers compose well defined service interfaces allowing solution components to reliably interoperate and be flexibly recombined and reused in new solutions.

Figure 7 illustrates our reference SOA.

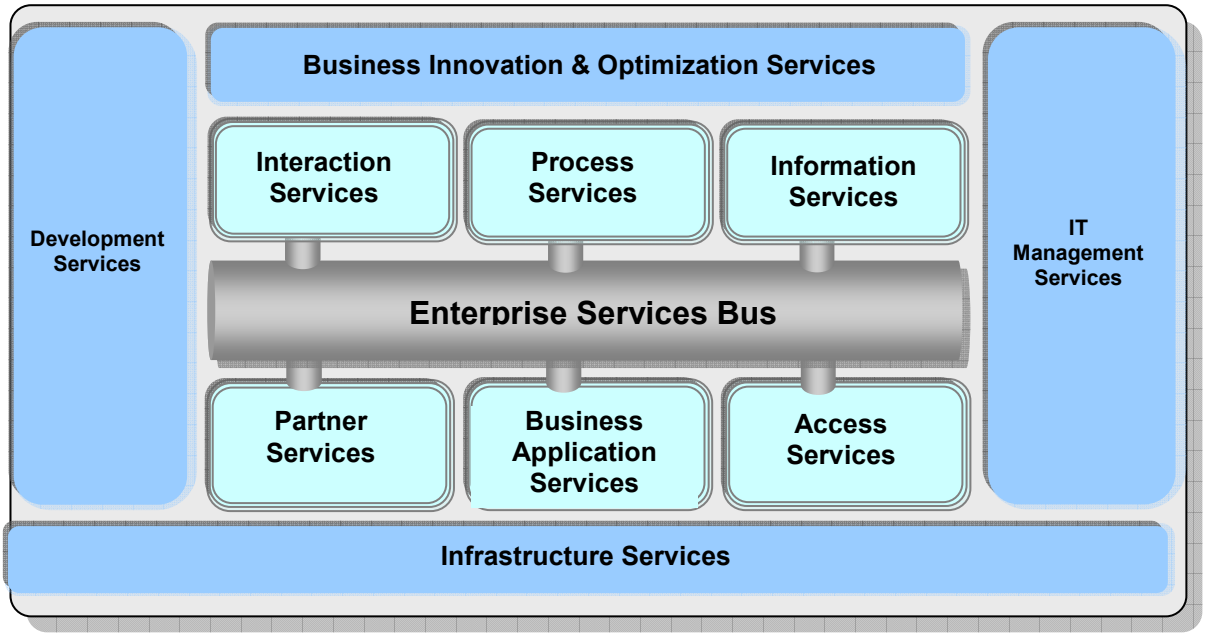


Figure 7. Reference Services Oriented Architecture

Below is a brief description of each component in the reference architecture:

- Development Services – An integrated environment for design and creation of solution assets.
- Business Innovation & Optimization Services - Facilitate better decision-making with real-time business information
- IT Management Service - Manage and secure services, applications and resources
- Infrastructure Services - Optimize throughput, availability and performance
- Interaction Services - Enable collaboration between people, processes & information
- Process Services - Orchestrate and automate business processes
- Information Services - Manages diverse data and content in a unified manner
- Partner Services - Connect with trading partners
- Business Application Services – Provide a robust, scaleable, and secure business applications and tools services environment
- Access Services - Facilitate interactions with existing information and application assets
- Enterprise Services Bus – Facilitate communication between services

The following table is a mapping of the Services components in a Janus solution, some of which were depicted earlier by Figure 7, onto the Services Oriented Architecture depicted above.

Services Category	Component Mapping
Business Innovations and Optimization Services	• Business process modeling and transformations
Development Services	• Data modeling tools • Meta data modeling tools
IT Management Service	• Application life-cycle management • Application orchestration and provisioning
Infrastructure Services	• Database and file management systems • Authentication and authorization services (e.g., CSM, LDAP) • Audit tracking services (e.g., CLM) • Process choreography services
Interaction Services	• Enterprise portal services • Collaboration services (e.g., annotations, email, blogs, chat, web meetings)
Process Services	• Application / Study tracking services (e.g., COMIS, DARRTS) • Application / Study review workflows
Information Services	• Janus repository • ArrayTrack and ECG Data Warehouse • Data validation services • Submission management • Data load services • Enterprise document management (e.g., EDR) • Application store management • Vocabulary services (e.g., EVS) • Application store management • Analysis scripts / programs (e.g., Open Toolbox tools scripts) • Data version management • Data model life-cycle management
Partner Services	• CRIX Clinical Submission data pre-validation services • CRIX Investigator Registry Services (Firebird)
Business Application Services	• Data analysis, review and reporting tools, e.g., SAS, I-Review, WebSDM, SPlus, Tox Vision
Access Services	• Application integration framework (e.g., Open Toolbox) • Data integration framework (e.g., CTOM) • Data federation services

4.4. Janus Deployment at FDA

Figure 8 depicts the future deployment architecture of Janus at FDA. The electronic submissions (e-CTD or non-eCTD) received from sponsors will be stored in a file server called an electronic document room (EDR). The tabulation datasets in SDTM (plus Define.XML) and SEND formats[2] included in FDA applications, which may consist of one or more studies, are then passed through a validation and load step from the EDR. The validation is done using the CDISC specifications for the syntax and FDA business rules to ascertain syntactic and semantic conformance. The errors detected by the data validation step, i.e., any syntactic and/or semantic violations, are monitored by FDA Systems Administrators who then trigger a response back to the sponsor on the status with an acceptance or a request for resubmission. The validated SDTM and SEND data and the associated meta-data (available through the sponsor-provided Define.xml file) are then placed in a staging area in a relational repository from which it is loaded into the Janus repository.

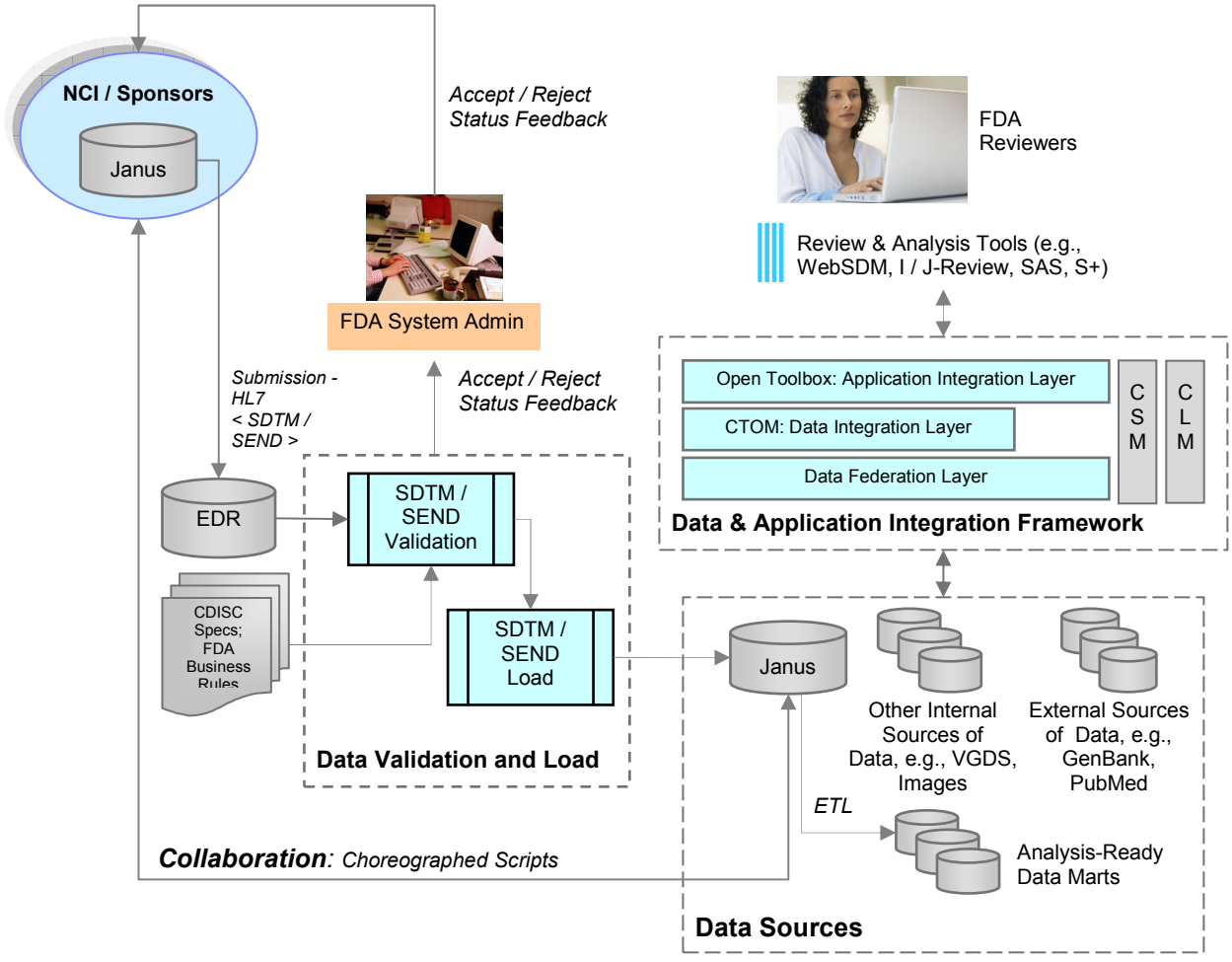


Figure 8. Envisioned Janus Deployment at FDA

Janus, being a normalized and extensible data model for efficient storage of SDTM and SEND datasets, is not ideally designed suited for study-specific or complex data mining scenarios. For such applications, one may pull data out of Janus into analysis ready data marts designed specifically for

[2] SDTM data is currently submitted as SAS Transport data files but there is an effort to encapsulate them in HL7 V3 messages.

fast query performance based on the analysis scenario.[3] Besides Janus, there are other data sources that a reviewer may access during the review process such as the Voluntary Genomic Data Submission (VGDS) repository ArrayTrack and external data sources such as those provided by NCBI's Entrez web service (GenBank, PubMed).

We envision a framework consisting of

1. a data federation layer that will create a dynamic virtual warehouse over a heterogeneous set of data sources (Janus, other intenal data sources like genomics and image repositories, external reference sources like GenBank and PubMed);
2. an object-oriented abstract data model based on caBIG™'s CTOM which will facilitate semantic data integration; and
3. an application integration framework based on the FDA's "Open Toolbox" initiative which will provide for reusability and interoperability of analysis modules between analytical applications.

A variety of review applications can be supported through this framework, e.g., WebSDM, I-Review/J-Review, SAS, S-Plus and ToxVision. During a review process, analytical programs and datasets need to be exchanged between the sponsors and FDA. We envision that the ADaM standard will support this interchange. Currently, Janus does not support mapping of ADaM data but there are hooks in place in its design allowing for future extensions as the standard becomes more mature by the industry.

4.5. Janus Deployment at NCI

Figure 9 depicts the envisioned Janus deployment at NCI. As indicated, NCI will leverage Janus for managing FDA submissions of SDTM and SEND data. In addition, NCI researchers will be able to use Janus as an analytical cross-trial data repository for retrospective and prospective clinical research.

Currently, NCI has all its clinical trial data from the Center for Cancer Research stored in C3D (Oracle Clinical based Clinical Data Management System) which will feed into an operational data store called CTDB.[4]. CTDB will also receive data from external cancer centers, cooperative groups and community clinical oncology programs (CCOPs) as part of the Cancer Therapy Evaluation Program (CTEP) and Division of Cancer Prevention (DCP) funded and sponsored clinical trials. Currently, the Clinical Trial Monitoring System (CTMS) and Clinical Data System (CDS) manage all clinical submissions from the cancer centers, cooperative groups and CCOPs into CTEP.

A subset of the CTDB data repository that is relevant for submission to FDA will then be loaded into NCI's own instance of Janus to facilitate:

1. submissions of INDs to FDA;
2. collaboration during the review process with FDA, via ADaM analysis datasets; and
3. collaboration in programs with joint clinical development partners such as biopharmaceutical companies or academic medical centers.

NCI researchers will also use the same data and application framework as the FDA reviewers, using a similar set of analytical tools such as SAS, I-Review/J-Review and WebSDM.

[3] Materialized SDTM views created on top of Janus is a very simple form of such a data mart.

[4] The Clinical Trial Working Group (CTWG) at NCI has proposed a vision to store all operational data in a centralized repository called the Clinical Trial Data Base (CTDB). This work is still in its preliminary stages.

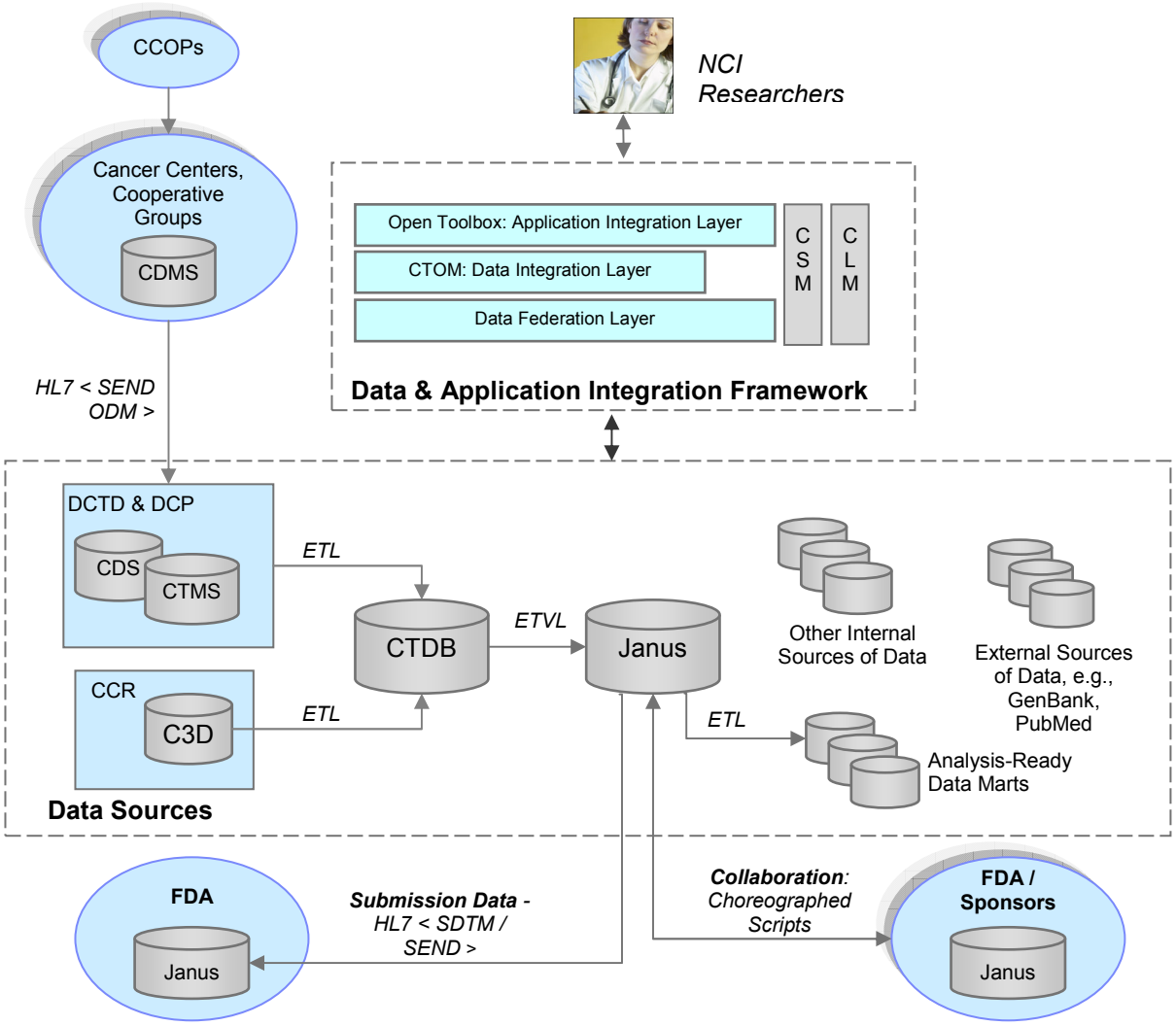


Figure 9. Envisioned Janus Deployment at NCI

4.6. Janus Deployment in the Biotechnology and Pharmaceutical Industry

Janus was created to enable effective collaboration environment between biopharmaceutical sponsors and FDA during the review process, expediting the tedious regulatory review and approval process. As shown in Figure 10, biopharmaceutical sponsors can deploy their own local instances of Janus that will contain all their tabulation datasets for the clinical and non-clinical submissions. The operational data generated during the execution of a trial, captured in electronic or paper case report forms, will be fed into the sponsor's CDMS data repository using the Operational Data Model (ODM) developed by CDISC.

SDTM datasets are generated from the CDMS environment using custom ETL scripts and then loaded into Janus along with the non-clinical data in SEND format. Biopharmaceutical sponsors will load the raw data from the case report forms and the derived data (such as computed outcome attributes) into Janus. Tabulation dataset submissions are directly prepared from Janus in SDTM and SEND format and included in an electronic Common Technical Document (eCTD) submission to FDA.

The regulatory developers in a biopharmaceutical company will use the same data and application integration framework as described in Section 4.4. Since the underlying data model (Janus) is the same at FDA and sponsors, analytical programs can be readily shared so that they can be executed at

each location via a point-and-click interface to replicate the same analysis environment, sharing analysis datasets and program in the ADaM format. This will significantly improve productivity and reduce the review time of applications – statistics show that some 80% of the time that takes to drive an application through the approval process is spent in preparing the data for analysis.

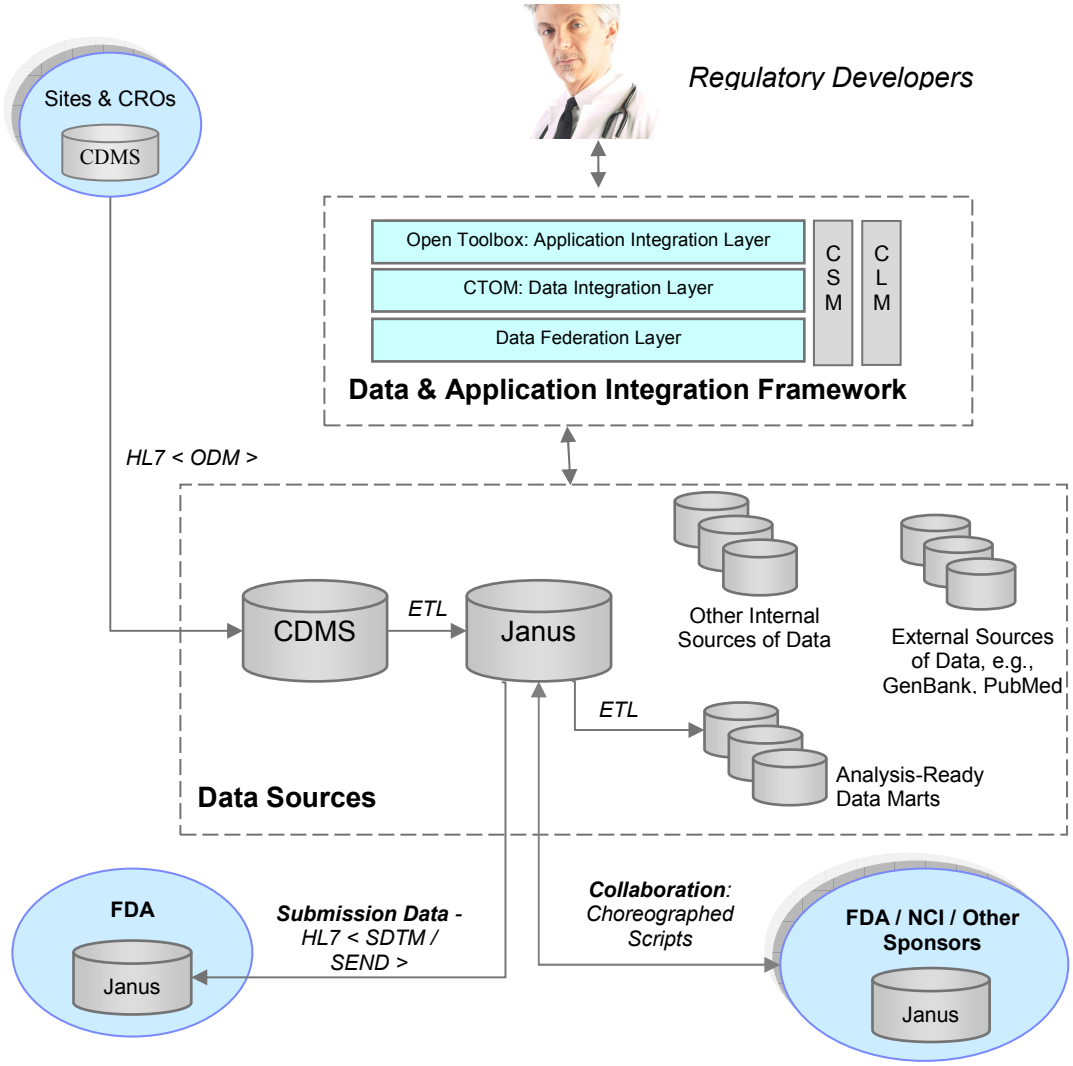


Figure 10. Envisioned Janus Deployment at Biotechnology and Pharmaceutical Sponsors

4.7. Features

The key features of the Janus solution are as follows:

- **Data Interchange using CDISC Data Standards:** Janus is an extensible and normalized relational data repository designed to store SDTM, SEND and eventually ADaM tabulation datasets. The SDTM and SEND tabulation datasets were designed, respectively, for the interchange of clinical and non-clinical data. The ADaM model was designed for the sharing of analysis datasets and programs between sponsors and FDA
- **Analysis-Ready Data Marts:** Janus is ideally designed for the efficient storage of tabulation datasets. The high performance required by complex data mining and study-specific analysis can be enabled by the creation of a variety of analysis-specific data marts, using star-schema models or simpler denormalized data formats such as materialized SDTM views.

- **Data Validation and Load:** A flexible approach to data validation is required, using CDISC syntactical specifications, sponsor-provided study-level metadata in Define.xml files, and FDA business rules, all of which may change over time. This module should be able to dynamically lookup the sponsor-provided study-level metadata and validation rules and implement study-level, table-level and row-level checks. It should also be able to generate error logs in a standard format to enable efficient collaboration between FDA and sponsors to streamline the rectification and resubmission process in case of rejection.
- **Data and Application Integration Framework:** For flexible and dynamic integration of heterogeneous data sources such as the Janus relational repository, ArrayTrack and external web services such as Entrez, a data federation layer is supported. NCI's caBIG™ initiative has proposed CTOM, an object-oriented abstraction model, for applications to access and interchange clinical trial data. This takes an implementation-neutral approach so that any underlying clinical trial data store can be mapped into CTOM to provide a unified access and data sharing environment for applications. FDA recently proposed an "Open Toolbox" concept that would allow interoperability and reusability of analysis modules among analytical applications. This would foster an enhanced analysis sharing environment between FDA and sponsors.
- **Security & Audit Compliance:** NCI's caBIG™ initiative has proposed a comprehensive security and audit compliance solution involving the Common Security Module (CSM) and Common Logging Module (CLM) to unify the implementation of security and auditing layer among applications. This supports authentication, authorization and standard user provisioning methods for managing the creation and modification of users and their access to data and applications and common logging and auditing capability for all applications.

5. Solution Rollout Plan

When implemented, Janus will be a transparent, yet essential, layer of a new environment in which clinical data can be submitted and reviewed for Investigational New Drugs (INDs), Biologic Licensing Agreements (BLAs), and New Drug Applications (NDAs). Users will interface with Janus through a metadata layer that communicates directly with the commercial analytical tools, such as SAS, WebSDM, I-Review, and others.

Key benefits with which Janus can provide stakeholders include the ability to:

- Automate cumbersome tasks through adoption of SDTM and SEND data standards involved in patient care and clinical studies.
- Enable cross-trial analysis for advanced and robust analysis for detecting clinical trends.
- Detect clinical trends more easily through visualization.
- Provide an integrated analysis platform for mining the existing volume of clinical data to test and validate clinical hypotheses.
- Facilitate effective collaboration (exchange of data, sharing of reproducible analysis and experiments) amongst biopharmaceutical companies, NCI, FDA, CROs, trial sites and other research partners.

FDA, NCI and other industry stakeholders should work together to establish the Janus Change Control Board (CCB). The JANUS CCB will be a board of representatives from the stakeholder community that will review suggested changes to the JANUS data model and ratify any official changes to the model.

The adoption of Janus across the industry is dependent on clinical data being available in SDTM standard, with syntactically and semantically robust content, so it can be loaded into the Janus repository easily, and then analyzed and submitted to the FDA. Adoption has been slow but is now gaining traction as a result of the FDA's Critical Path Initiative and SDTM guidance.

Increasing SDTM adoption will require influential players like the FDA, tool vendors, and the industry itself to take a more direct role. In a similar vein, one would expect eventual effort from FDA in issuing guidance on the adoption of the SEND standard for non-clinical data as well. The following table outlines how various members of the constituency may affect the adoption rate of the SDTM standard.

The increased adoption of this particular standard is of critical value to the realization of our vision and widespread deployment of the Janus solution.

Constituency	Actions
FDA	▪ Develop a regulation that requires SDTM for all electronic submissions. This would be the most effective method, but may take years to be realized, given the process that is in place for finalizing regulations. ▪ Revise the guidance on SDTM data format submissions. The FDA has already received several submissions that included SDTM data based on the official FDA guidance of the e-Common Technical Document, but data quality has been inconsistent. The FDA should work with CDISC to improve the instructions and develop training materials to increase consistent adherence for the industry leaders who have taken these initial steps. ▪ During pivotal meetings between FDA and sponsors, e.g., those at the end of Phase I and Phase II, FDA reviewers can encourage sponsors to submit data in SDTM format. ▪ Use incentives like earlier completion of reviews. This may require congressional action to change the PDUFA (Prescription Drug User Fee Act) law. The current PDUFA law expires in 2008. FDA and the biopharmaceutical industry are working on the new enactment, which offers a timely opportunity for mandating adoption of the SDTM standard. ▪ Use existing public forums to announce and discuss preference for CDISC standardized data. ▪ Conduct a formal Return on Investment (ROI) assessment using a balanced scorecard approach. This includes devising metrics that concretely model the added business benefits in terms of increased productivity of the development, submission, and review of a new medical product application. A clear business case with a formal ROI analysis and measures will certainly influence the adoption of SDTM and Janus in a positively.
Vendors	▪ Begin integrating SDTM functionality into their products to support the industry.
Sponsors	▪ Incorporate SDTM into their eCRF, e-protocols, and all clinical data prepared for internal analysis and submissions. This will foster information reuse and avoids transformations errors which will ultimately improve data quality. ✓ Target biopharmaceutical leaders who are already participating in CDISC consortia and other groups that rely on standardized data to achieve their goals. ✓ Target medium sized biopharmaceutical companies who have not yet adopted any standards. Such companies are often more receptive to change regarding technology and organizational shifts. ✓ Demonstrate the benefits of collaborating with current partners, such as CROs, but also of expanding the base of partners. ✓ Demonstrate the benefits to facilitate the ability to reuse data for new indications and multiple submissions for an application, answering review questions faster, and addressing safety issues by pooling data.

5.1. Rollout at FDA

FDA's interest in Janus is multi-faceted. The FDA has more human and animal research data than anyone else in the world, but currently data submitted to the FDA is in different formats and not easily accessible. Receiving data in SDTM format is a critical first step in reducing the administrative burden on reviewers and others who prepare the data for review. The FDA is therefore moving forward and setting the example for the industry.

Below are some the key **long-term** goals of FDA, based on the interviews conducted with some of the key individuals at FDA:

- Create division-specific panels of pre-set safety analyses that run on SDTM data to best suit users' needs. Pre-set analyses can then be run as a batch and then applied to a Safety Review Works template. Having this capability would allow Medical Officers or reviewers to use their expertise and focus on providing interpretations for the analysis report.
- Enable reviewers to conduct analysis, share their experiments with sponsors, and corroborate results instead of the current practice of asking sponsors to conduct analysis on their behalf. Putting such an effective collaboration platform in place will result in saving time for decision making on approvability, review, and labeling.
- Establish a central repository for failed trials, knowledge learned from placebo arm, and successful submissions to explore lessons learned based on the wealth of information from submitted data.
- Leverage Janus to rebuild the research culture at FDA. With a syntactically and semantically robust content, Janus will provide an appealing platform that fosters collaboration and research-oriented activities.
- Establish Open Toolbox, a recent initiative by FDA, as an application integration platform that promotes development of user-friendly tools, reusability and interoperability of various analysis modules. The same vision is stressed in FDA's Critical Path initiative.

In the short term, however, successful rollout of Janus at FDA involves:

- Integrating it with the current commonly used analytical tools. Work is underway to integrate Janus with three commercially available analytical tools that are used by reviewers in the FDA: SAS, WebSDM, and I-Review.
- Heightening awareness to internal stakeholders to reduce resistance. Communications need to include a broad message addressing the benefits of Janus and analytical tools addressing the daily struggles of the target audience and support of the Critical Path.
- Rolling user-friendly tools to target audiences, and providing appropriate level of training based on immediate needs and timely opportunity to apply training to immediate responsibilities. Based on our recent interviews within the FDA, we categorize the target audience into the following three categories:
    1. Statisticians and medical reviewers who are typically power-users and conduct statistical analysis using tools like SAS and I-Review;
    2. Medical reviewers who typically rely on other staff to conduct complex analysis for them and only use pre-canned reports and simple tools like Patient Profile Viewer for cross-trial analysis; and
    3. Clinical pharmacologists who use modeling and simulation tools to analyze pharmacokinetic data in phase I, II, and III studies to better understand toxicology of compounds.
- Creating a governance group made up of key stakeholder representatives from the affected offices. This group would represent the affected divisions and address needs of clinical users and technical issues. They would have a two-way communication with the Janus change control board that would have direct responsibility of applying technical updates.
- Conducting an effective training program. Clearly, training should first target those who will be immediately affected so that they will be better prepared to cope with change. A well-designed training program relates the training material with everyday responsibilities, and accommodates a "Just in Time" component that makes it available to users when they need it

the most, e.g., when submissions are coming. A "train the trainer" approach will make the training program mode scalable, allowing it to be conducted with relatively small staff.

- Leveraging Ambassadors, knowledgeable peers and role models who are often self-appointed champions who can help heighten awareness and reduce resistance to the adoption of new solutions. They are typically influential individuals in their own organizations, and in some cases, can even impact other divisions. Examples might include representatives from the Office of Biostatistics and Office of New Drugs.

5.2. Rollout at NCI

The NCI conducts research trials internally, and works with many partners including medical centers and biopharmaceutical sponsors. Having a robust data platform that enables complex data analysis scenarios and effective collaboration is essential to a vibrant research organization. Due to lack of widespread use of standards, controlled vocabularies, and common analysis tools, however, NCI faces the same challenge as many other organizations, a systemic lack of collaboration culture.

Having a clinical trial database built on a Janus data model, again with syntactically and semantically robust content, will:

- Enable researchers to conduct analysis, and share their experiments in a reproducible fashion with other researchers within and outside of the organization.
- Facilitate a research feedback loop through data mining across studies to help researchers develop and prioritize new studies. Having broader access to toxicity and efficacy data will help plan new trials or compare results to determine if results of an ongoing study are a normal or abnormal signal.
- Establish an environment that is easy to query and develop standardize reports across all studies to simplify reporting from the field to the NCI and beyond.
- Facilitate contributions to submissions of INDs, BLAs, or NDAs.
- Provide necessary data on trials for groups such as the Investigational Drug Steering Committee to make decisions and prioritize studies.
- Provide access to data across all NCI partners.
- Enable the integration of clinical data with other standardized data such as biomarker data.

5.3. Rollout at Biopharmaceuticals Companies

The current emphasis by FDA on a standard format for electronic clinical trial submission (SDTM / SEND), and stronger regulatory mandates that are anticipated in the coming years (2008-2010), are good indicators that such standards will be more universally adopted by sponsors in the near future. Currently, however, adoption is not yet widespread. We certainly observe movements in the industry where sponsors are experimenting with small projects using the emerging standards, assessing the impact on their infrastructure, and evaluating options for incorporating them into their business workflows.

With a stated intent to adopt Janus as the common repository for clinical trial submission data, and the analysis platform of choice, FDA is also sending very strong signals to the industry: having a common analysis platform as FDA in their own shop, and being able to reproduce the experiments and analysis results of regulatory reviewers effectively, will offer them the potential for reducing the approval cycle time. The net will be a clear benefit to all stakeholders, and ultimately, the pubic at large who will have access to safe and effective drugs in a timely manner.

6. Key Related Initiatives

Initiative	Relation to Janus	Comments
FDA Critical Path Initiative	Development of new scientific and technical tools for assessing safety and efficacy of new medical products with higher predictability, and improve efficiency and quality control of the manufacturing process	Janus will advance the objectives of FDA's Critical Path Initiative by making syntactically and semantically robust operational and summarized clinical data available to the entire community of user constituency (FDA regulatory reviewers, biopharmaceutical developers, and NCI researchers). It also enables effective communications amongst the user community expediting the currently tedious regulatory approval process and multi-institutional clinical trial research effort.
NCI Clinical Trial Working Group Report	Create a comprehensive database containing information on all NCI-funded clinical trials to facilitate better planning and management across clinical trial venues.	Janus will be a critical component of the comprehensive clinical trial database (CTDB) to help increase cooperation between NCI, FDA, and industry to enhance the focus and efficiency of oncology drug development.
Clinical Data Interchange Standards Consortium (CDISC)	Syntax and semantic standard for clinical data (ODM / SDTM), non clinical data (SEND), and Analysis Data (ADaM) interchange	Janus will adopt ODM. SDTM, SEND, and ADaM standards for data exchange with other sources of data, e.g., between instances of Janus, between Janus and operational data stores. Janus will adopt ADaM spec for capturing and sharing the review and analysis steps and results, providing an effective means of collaboration amongst FDA reviewers, sponsors, and NCI researchers.
CDISC/NCI Biomedical Research Integrated Domain Group (BRIDG)	Semantic standard for clinical trial data	Janus will leverage BRIDG to provide a harmonized repository of summarized clinical trial data enabling cross-clinical trial analysis.
NCI Clinical Trial Object Model (CTOM)	Data integration framework	Janus will leverage CTOM, a common object model for clinical trial data, to provide a data integration framework, enabling access to suite of independently developed data sources
FDA Open Toolbox	Data interchange specification	Janus will leverage the emerging Open Toolbox work intended to facilitate reusability and interoperability of analysis modules between analytical applications.
HL7	Data interchange specification	Janus will utilize HL7 standards that enable a common data interchange transport for the exchange of clinical trial data (*e.g.*, SDTM, SEND).
American Medical Informatics Association (AMIA)	The goal of the Global Trial Bank (GTB) is to help propel the development of a world-wide peer-reviewed repository of protocols and results from clinical trials of all types	The Janus data model could be leveraged to support the GTB vision of implementing a comprehensive database of clinical trial results including raw subject level data

7. Appendix

7.1. Janus Data Model

Figure 11 depicts Janus model's entities, the parent (Subjects) and the 3 children tables (Events, Findings, and Interventions), which correspond to their counterpart SDTM domain classes. In addition, the Janus model includes entities like Studies, Arm, Sites, Visits, Elements and others that support the SDTM trial design domains. The Qualifiers and Comments entities support the storage of special purpose domains such as Demographics and Comments. The string data in the domains belonging to three general classes is mapped to Qualifiers if it is a coded term and Comments if it is not coded. The Supplemental Qualifiers domain in SDTM that extends the schema of the general domains classes is also mapped into the Qualifiers and Comments table. Janus supports storage of controlled vocabulary through entities such as Codelists, Codes and Decodes.

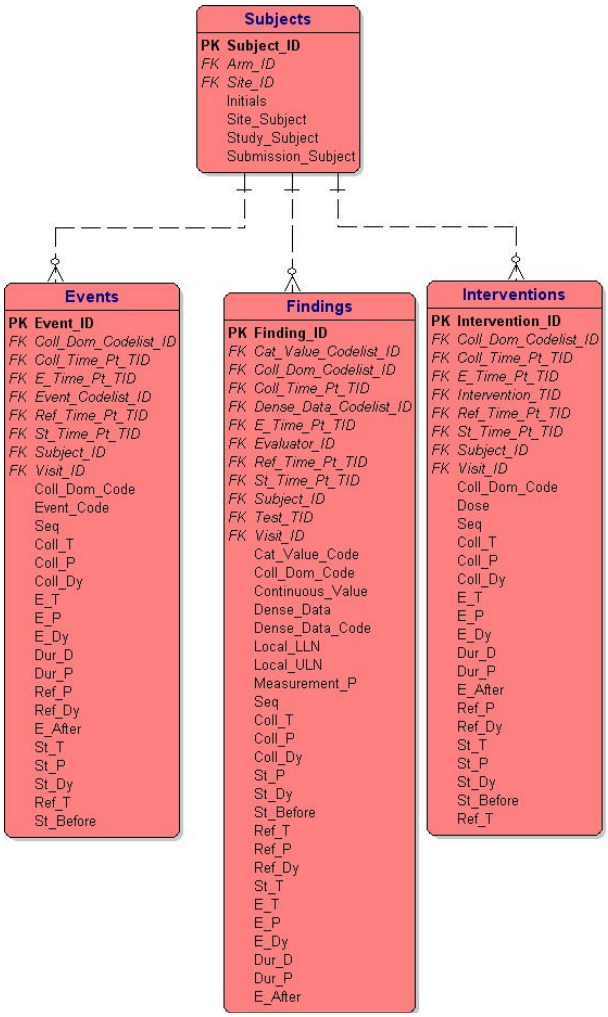


Figure 11. Key entities in the Janus relational data model

8. Glossary

ADaM	Analysis Data Model (http://www.cdisc.org/standards/)
BRIDG	Biomedical Research Interchange Domain Group (http://www.bridgproject.org/)
CCOP	Community Clinical Oncology Program
CCR	Center for Cancer Research
CDISC	Clinical Data Interchange Standards Consortium (http://www.cdisc.org)
CDS	Clinical Data System
COMIS	Center-wide Oracle Management Information System
CTDB	Clinical Trial Data Base
CTMS	Clinical Trial Management System
DARRTS	Document Archiving, Reporting and Regulatory Tracking System. An integrated system envisioned to house many of the FDA's core tracking systems.
DCP	Division of Cancer Prevention
DCTD	Division of Cancer Treatment and Diagnosis
ODM	Operational Data Model (http://www.cdisc.org/standards/)
CTOM	Clinical Trial Object Model
ETVL	Extract, Transform, Validate and Load – generally referring to tools for extracting data from one or more sources, transform it into another format, validate using data specifications and business rules and load it into a target data source
SDTM	Study Data Tabulation Model (http://www.cdisc.org/standards/)
SEND	Standard for Exchange of Nonclinical Data (http://www.cdisc.org/standards/)

9. References

R1. "Innovation or Stagnation; Challenge and Opportunity on the Critical Path to New Medical Products," U.S. Department of Health and Welfare, Food and Drug Administration, March 2004.

R2. "Restructuring the National Cancer Clinical Trials Enterprise." Report of the Clinical Trials Working Group of the National Cancer Advisory Board, National Cancer Institute, June 2005.

R3. "IBM's SOA Foundation – An Architectural Introduction and Overview," Version 1.0, November 2005 (http://download.boulder.ibm.com/ibmdl/pub/software/dw/webservices/ws-soa-whitepaper.pdf).

R4. "Clinical Trials Object Model (CTOM), Model Definition Report." Submitted to National Cancer Institute, Center for Bioinformatics by SCENPRO, 101 West Renner Road, Suite 130, Richardson, TX 75082.

R5. "The BRIDG model: a shared domain analysis model of regulated clinical research." Douglas B. Fridsma, MD, PhD, Center for Biomedical Informatics, Center for Pathology and Oncology Informatics, University of Pittsburgh.

R6. "caGrid: design and implementation of the core architecture of the cancer Biomedical Informatics Grid." Joel Saltz, Scott Oster, Shannon Hastings, Stephen Langella, Tahsin Kurc, William Sanchez, Manav Kher, Arumani Manisundaram, Krishnakant Shanbhag, and Peter Covitz, Bioinformatics, Oxford Journals, June 9, 2006.

R7. "caCore 3.1 Technical Guide," National Cancer Institute, Center for Biointormatics, May 10, 2006.